ARTICLE IN PRESS

Int. J. Human-Computer Studies ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Sonification of guidance data during road crossing for people with visual impairments or blindness ☆

Sergio Mascetti [a,b], Lorenzo Picinali [c], Andrea Gerino [a,b,*], Dragan Ahmetovic [a], Cristian Bernareggi [a,b]

[a] Universita degli Studi di Milano, Department of Computer Science, Via Comelico 39, 20135 Milano, Italy
[b] EveryWare Technologies, Via Comelico 39, 20135 Milano, Italy
[c] Dyson School of Design Engineering, Imperial College London, South Kensington, SW7 2AZ, London, UK

ARTICLE INFO

Article history:
Received 29 January 2015
Received in revised form
13 June 2015
Accepted 13 August 2015

Keywords:
Blind people
People with visual impairments or blindness
Sonification of guidance data
Spatial cognition

ABSTRACT

In the last years several solutions were proposed to support people with visual impairments or blindness during road crossing. These solutions focus on computer vision techniques for recognizing pedestrian crosswalks and computing their relative position from the user. Instead, this contribution addresses a different problem; the design of an auditory interface that can effectively guide the user during road crossing. Two original auditory guiding modes based on data sonification are presented and compared with a guiding mode based on speech messages.

Experimental evaluation shows that there is no guiding mode that is best suited for all test subjects. The average time to align and cross is not significantly different among the three guiding modes, and test subjects distribute their preferences for the best guiding mode almost uniformly among the three solutions. From the experiments it also emerges that higher effort is necessary for decoding the sonified instructions if compared to the speech instructions, and that test subjects require frequent 'hints' (in the form of speech messages). Despite this, more than 2/3 of test subjects prefer one of the two guiding modes based on sonification. There are two main reasons for this: firstly, with speech messages it is harder to hear the sound of the environment, and secondly sonified messages convey information about the "quantity" of the expected movement.

1. Introduction

Mobile devices provide new exciting opportunities for people with Visual Impairments or Blindness (VIB). Indeed, most commercial devices (e.g., based on iOS and Android) are accessible to people with VIB.[1] On one hand, this allows people with VIB to use most of the applications available on mobile devices, such as web browsers and email clients. On the other hand, accessible mobile devices can be used to implement assistive technologies, with great advantages for both developers and users. The developers can rely on well known platforms, for which there is plenty of documentation and software libraries, and which provide high level OS APIs to support accessibility (e.g., text-to-speech functionalities on iOS). For the final user, a single device capable of providing different assistive tools is cheaper, quicker to learn and more convenient (in terms of weight to carry, devices to charge, etc…).

Mobile devices also have two main advantages with respect to traditional ones (i.e., desktops and laptops). Firstly, they can be used on the move, hence can provide support in many situations in which it is impractical to rely on a traditional device. Secondly, mobile devices are equipped with hardware sensors such as GPS receivers, accelerometers, and gyroscopes, that can be used to acquire information about the user's context and position. In this context, it is not surprising that several research contributions in the last years focused on mobile assistive technologies. In particular, a number of solutions have been proposed to support autonomous mobility, for example by recognizing objects in the environment and notifying the user accordingly.

In this contribution we take into account the problem of guiding the user towards and over a zebra crossing (i.e., a particular type of pedestrian crosswalk also called "continental crosswalk" in the United States). This problem involves non-trivial computer vision techniques to recognize the zebra crossing pattern, as well as advanced spatial reasoning, based also on accelerometer data, to reconstruct the position of the crosswalk with respect to the user. In

☆This paper has been recommended for acceptance by E. Motta

* Corresponding author at: Universita degli Studi di Milano, Department of Computer Science, Via Comelico 39, 20135 Milano, Italy. Tel.: +39 02 50316219; fax: +39 02 87153726.
E-mail address: andrea.gerino@unimi.it (A. Gerino).

[1] In case the reader is unfamiliar with accessibility tools for visually impaired individuals, we suggest a short video introducing the main ideas http://www.goo.gl/mEI6Uz

http://dx.doi.org/10.1016/j.ijhcs.2015.08.003
1071-5819/

our previous work we describe the 'recognition' procedure used to identify the crosswalk and compute its relative position (Ahmetovic et al., 2014).

Other existing contributions in the field focus on the recognition procedure (Se, 2000; Uddin and Shioyama, 2005a, 2005b; Ivanchenko et al., 2008, 2009; Ahmetovic et al., 2011). However, a different challenge is now arising: how to guide the user employing audio instructions. Two contrasting objectives emerge. On one hand audio instructions should provide precise and responsive information. On the other hand, they should not distract the user's attention from the surrounding environment.

This paper presents two auditory guiding modes based on data sonification. The two guiding modes are similar, with the main difference being that one produces mono sound (i.e., one single sound signal) and the other produces stereo sound (i.e., two different sound signals, one for the left and one for the right ear). From the applicative point of view, a major difference can be noted; stereo sonification requires the user to wear headphones, while mono sonification can also be reproduced from the device's internal speaker.

The sound design process was conducted employing a user-centric approach, frequently considering end users feedback and carrying out a preliminary evaluation session. The two sonifications, together with a guiding mode based on speech messages, have been implemented in the *ZebraX* prototype, an iPhone application that adopts a state-of-the-art algorithm to detect zebra crossings. *ZebraX* was then used to conduct three sets of evaluations aimed at assessing the effectiveness of the guiding modes. Experimental results show that the three guiding modes can effectively support the test subject to align with the zebra crossing and to actually cross it. Still, the two guiding modes based on sonification are less immediate to use, and some subjects required frequent hints (in the form of speech messages) to correctly interpret the sonified instructions.

Despite this, two results are available supporting the applicability of the two guiding modes based on sonification. Firstly, after a few minutes of training only, there is not a statistically significant difference in the performance (e.g., crossing time) between the three guiding modes. Secondly, 75% of the subjects declared that they preferred the two guiding modes based on sonification. Furthermore, they reported that hearing sounds from the surrounding environment, a very important task when crossing a road, is more difficult with the speech mode than with the two sonifications.

Section 2 describes the related work as well as the system architecture of *ZebraX*. The three auditory guiding modes are presented in Section 3 while Sections 4 and 5 present the results of two evaluation sessions. Section 6 concludes the paper and highlights future work.

2. Background

It is well known that independent mobility is very challenging for people with VIB. Blind people can find their way by means of a white cane or a guide dog, whereas partially sighted people can also rely on their residual sight. The main difficulties are related with avoiding obstacles along the way (e.g., people on the sidewalk, trash bins, poles, etc.), finding a target (e.g., stairs, doors, intersections, etc.) and getting information reported on pedestrian signs (e.g., crossing a road over a zebra crossing when the traffic light is green, etc.).

Over the years, many solutions for supporting independent mobility have been investigated in scientific literature. In particular, in the following paragraphs we report the main findings in the field of pedestrian crosswalk detection (Section 2.1) and guidance (Section 2.2). In both cases, we focus our attention on the technique to convey information to users with VIB.

In more recent years, commercial applications for orientation and mobility of people with sight impairment became available as well. We briefly describe a few of them in Section 2.3. Finally, in Section 2.4 we describe the architecture of the *ZebraX* application.

2.1. Solutions for pedestrian crossing

In 2000, Stephen Se proposed the first technique to recognize pedestrian crosswalks with the goal of supporting people with VIB (Se, 2000). The main limitation of this solution is that it fails to recognize a zebra crossing when its pattern is not completely in the camera field of view, or when it is covered by an object (e.g., a car). Uddin et al. address this problem and propose a solution to improve the effectiveness of the detection algorithm through bipolarity feature check and projective invariant (Uddin and Shioyama, 2005a, 2005b). These first contributions focus on the computer vision algorithm, and do not address the problem of how to interact with the user.

Successively, Ivanchenko et al. illustrate two techniques for detecting pedestrian crosswalks through the camera of a smartphone. The first technique focuses on zebra crossing and describes an application that produces an audio tone each time a zebra crossing is recognized (Ivanchenko et al., 2008). An experimental evaluation with two blind test subjects is presented to assess the ability of an individual to determine whether or not there is a crosswalk at a traffic intersection. The results shows that both test subjects were able to find the zebra crossing in each one of the 15 trials. The second technique is aimed at recognizing United States transverse crosswalks (also known as 'two stripes' crosswalks) (Ivanchenko et al., 2009). In this solution, the recognition algorithm also detects lateral shift of the person with respect to the two-stripes crosswalk. The presence of the crosswalk is signaled with a short low-pitched tone, followed by a high-pitched tone. If only one single stripe is detected, a low-pitched tone is emitted. If the second stripe is detected later, and the first one is still in the filed of view, a high-pitched tone is emitted. No sound is generated if no stripe is detected. After detecting the two-stripes crosswalk, the application reproduces a speech message reporting the position of the person (i.e., inside, on the left or on the right of the crosswalk). An experimental evaluation conducted with two blind test subjects shows that individuals are able to find the crosswalk and are aware of their position with respect to the crosswalk in six cases out of eight trials.

The two solutions proposed by Ivanchenko et al. are extended by Ahmetovic et al. who, focusing on zebra crossings, propose a technique to compute a set of 9 qualitative relative positions of the user with respect to the crosswalk, each one corresponding to a user action (e.g., go ahead, rotate right, step left, etc., Ahmetovic et al., 2011). These actions are conveyed to the user in the form of speech messages and can guide the person to the best crossing point (i.e., in the middle of the first stripe). A qualitative experimental evaluation was conducted with five blind test subjects. These were required to complete two tasks. The first one involved the simple detection of a crosswalk positioned in front of the users. The second one involved the detection and location of a crosswalk on the sides, followed by crossing the road. All test subjects successfully accomplished the first task, while all users except one accomplished the second one. Two test subjects reported that a sound-based message would convey information more promptly.

In 2014, Ahmetovic et al. proposed the *ZebraRecognizer* algorithm to recognize zebra crossing (Ahmetovic et al., 2014). The algorithm rectifies the ground plane, hence removing the projection distortion of the zebra crossing features. This allows to compute the position of the user with respect to the crosswalk, producing